

# Computational Identification of Novel Missense Variants in Coding Regions of Genes Associated with Intellectual Disability

Anna Chang Liu<sup>1</sup>, Junmeng Yang<sup>2</sup>, Tina Yuan<sup>3</sup>, Yongsheng Bai<sup>4\*</sup>

<sup>1</sup> Appleby College, 540 Lakeshore Rd West, Oakville, Ontario, L6K 3P1, Canada.

<sup>2</sup> Shanghai Starriver Bilingual School, Minhang, 201100, Shanghai, China.

<sup>3</sup> The Roeper School, 1051 Oakland Avenue, Birmingham, Michigan, 48009, USA.

<sup>4</sup> Next-Gen Intelligent Science Training, Ann Arbor, Michigan, 48105, USA.

\* Corresponding author. Tel.: 1-817-559-3448; email: bioinformaticsresearchtomorrow@gmail.com

Manuscript submitted November 28, 2020; accepted February 11, 2021.

doi: 10.17706/ijbbb.2021.11.2.22-33

---

**Abstract:** Single nucleotide variants (SNVs) are mutations in the DNA sequence of a gene that are responsible for a variety of genetic disorders. Exonic variants located in the coding region of a gene can be especially harmful to an organism when they alter the biological functions of proteins. In this study, we performed multiple functional annotations and sequence alignments to assess the impact of SNVs in coding regions of genes associated with intellectual disability (ID). Through bioinformatics analysis, we identified three significant motifs in the coding region of *ACTB*, *PTPN11*, and *GNAO1* prioritized by our pipeline. Detailed examination of these three genes revealed that eight variant coordinates were located in the three motifs, two of which from *ACTB* are reported to be pathogenic by MISCAST and three from *GNAO1* are reported likely pathogenic according to clinical significance from National Center for Biotechnology Information (NCBI). In particular, we discovered two novel SNVs located at chr16:56370698 and chr16:56370711 in *GNAO1* that are highly likely to be pathogenic and possibly associated with ID. Overall, our results revealed the pathogenic roles of SNVs within coding regions of ID candidate genes. Future work is to automate this process through developing a bioinformatics software that can identify novel SNVs in coding regions responsible for various human diseases.

**Key words:** Coding region, intellectual disability, protein function, single nucleotide variants.

---

## 1. Introduction

Intellectual disability (ID) is a common neurodevelopmental disorder with problematic intellectual functioning and adaptive behaviour which significantly limit one's abilities to perform everyday social and practical skills [1]. One can be diagnosed with this condition prior to the age of 18 if the following criteria are met: 1) deficits in intellectual abilities such as "learning, problem solving, and judgement" confirmed by exams performed by doctors and individualized standard IQ testing scores lower than 75; 2) deficits in adaptive functioning such as "communication and independent living" which significantly hamper one's capability to meet their social responsibilities; 3) Delays in skill development (language and motor skills) compared to the average age [2]. Intellectual disability can appear as mild, moderate, or severe. Mild levels of ID are hard to detect until later in school age when a child may have difficulty with academics and suffer from recurring mental health issues [2].

Genetic brain disorders are caused by mutations in a gene, either from inheritance, environmental exposures, or both [3]. Previous studies have identified many ID-related genes. For instance, 26 orphan receptors found in the rhodopsin (class A) family of G protein-coupled receptors (GPCR) are shown to be extensively expressed in the mammalian brain and highly involved in neurodegenerative diseases and psychiatric disorders including Alzheimer's disease, Parkinson's disease, neuroinflammation, inflammatory pain, bipolar and schizophrenic disorders, epilepsy, anxiety, and depression [4], [5]. People with significant brain disorders fail to produce enough of a certain protein or produce malfunctioning proteins as a result of variations within their DNA sequences [3].

Existing microarray studies and exome sequencing studies have demonstrated the importance of *de novo* copy number variants (CNVs) as well as single-nucleotide variants (SNVs) in causing intellectual disability [6]. In a clinical experiment conducted by scientists from Radboud University Medical Center, a whole-genome sequencing was applied to 50 patients with severe ID and their unaffected parents. Their results showed 84 *de novo* SNVs altering the coding region (CDS), which revealed a "statistically significant enrichment of loss-of-function mutations as well as an enrichment for genes previously implicated in ID-related disorders" [6]. In fact, many other studies have also identified *de novo* variants harbored from ID related genes specifically within the CDS, a section of the DNA sequence responsible for encoding proteins [7], [8]. This correlation suggests that SNVs altering the CDS of a gene are major contributors to ID.

Nevertheless, given the genetic heterogeneity of ID and its relationship with the constantly changing environment, newly identified genes and respective mutations account for only a small proportion of ID cases [9]. Therefore, an automated and effective analysis pipeline of genome sequencing is necessary to effectively identify and prioritize the causative variants within ID candidate genes.

In this research project, we conducted bioinformatics analysis to identify novel ID-related SNVs exclusively in the coding region harbored in genes associated with ID. Other crucial evaluations such as functional annotation, gene tissue expression, conservation across species, conserved domain analysis, protein-protein interactions (PPI), and protein structure analysis were also performed to provide more insights into the target group of genes and their functions.

## **2. Materials and Methods**

### **2.1. Selection of Candidate ID Genes**

To obtain the initial list of genes related to ID, we collected known genes from existing studies published on reputable science journals and databases (e.g. PubMed). The selection criteria we followed were: 1) genes were all targeted by microRNAs and 2) published papers are from the past ten years and related to ID studies and clinical evaluations. We then manually screened out aliases and replaced them with their official gene names as suggested by The National Center for Biotechnology Information (NCBI) ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). In total, we reviewed nine published papers and collected a list of initial genes for subsequent analysis [10]-[16].

### **2.2. Functional Annotation and Tissue Expression**

The list of ID candidate genes were processed by the Database for Annotation, Visualization and Integrated Discovery (DAVID) ([david.ncifcrf.gov](http://david.ncifcrf.gov)) to identify their enriched biological themes, particularly Gene Ontology (GO) terms, of the genes and categorize them to respective functional-related gene groups. We limited our search to genes with the GO term 'neuron' from the top ten clusters and assessed their gene expression, especially in brain tissues, by checking them against the Genotype-Tissue Expression (GTEx) database ([gtexportal.org/home](http://gtexportal.org/home)) and the University of California Santa Cruz (UCSC) Genome Browser

(genome.ucsc.edu). Only genes that are highly or relatively highly expressed in brain tissues compared to the other genes were selected for performing protein sequence alignments.

### **2.3. Multiple Protein Sequence Alignment and Conserved Domain Analysis**

The purpose of protein sequence alignment is to find conserved domains within the ID candidate genes and evaluate their association with brain disorders. We retrieved protein sequences for the selected genes from Ensembl Biomart (ensembl.org) and fed them through the Multiple Sequence Comparison by Log-Expectation (MUSCLE) (www.ebi.ac.uk/Tools/msa/muscle) to generate alignment results. We viewed the results using Jalview (www.jalview.org) that clearly showcased the conservation value and qualities of the many domains. Through examining the results, we excluded *TUBB2A* from further analysis because it did not show similar conserved domains as the other genes.

Furthermore, we ran NCBI Batch CD-Search tool (ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) for the genes (not including *TUBB2A*) with conserved blocks in their protein sequences and identified their super-families. Within these families, we ran STRING functional protein association network (string-db.org) to inspect protein-protein interactions (PPI) between these genes. Lastly, we also explored homology and the conservation of these domains in other species (e.g. mouse and rat) at protein level using Protein Basic Local Alignment Search Tool (BLASTP) (blast.ncbi.nlm.nih.gov) and found that all selected genes have orthologs in mice and rats. Genes that showcased clear conserved domains and conservation across other species were selected for motif identification.

### **2.4. Phylogeny Construction for ID Candidate Genes Highly Expressed in Brain**

In order to build the phylogenetic tree, we took the MUSCLE alignment file and fed it to the Molecular Evolutionary Genetics Analysis (MEGAX) (www.megasoftware.net) and generated the Maximum Likelihood (ML) tree. We adopted the default tree-building parameters for ML Heuristic Method such as: uniform rates, all sites, Nearest-Neighbor-interchange (NN), no branch swap filter, and 4 threads in total.

### **2.5. Motif Identification and dbNSFP Database Variant Functional Prediction**

We used Multiple Expression motifs for Motif Elicitation (MEME) Suite (meme-suite.org) to identify the motifs among the CDS sequences of each of the ID candidate genes with default parameters and inspected whether genes enriched with identified motifs containing SNVs could contribute to ID. Although there are many databases available, we chose to use the database for Nonsynonymous SNPs' Functional Predictions (dbNSFP) to predict functions for SNVs exclusively in the CDS. This database includes more than 300 unique functional annotations for only nonsynonymous SNVs (nsSNVs change the final protein product) in the human genome [17]. Then, we examined if any of dbNSFP reported variants are located in the motifs identified by MEME.

### **2.6. Verification of the Nature of Identified Variants**

In total, we obtained three genes (*ACTB*, *GNAO1*, and *PTPN11*) from motif identification result of genes with conserved domains and high expressions in brain tissues. We then prioritized this list of genes harboring nsSNVs based on their damaging significance reported by dbNSFP and ClinVar (ncbi.nlm.nih.gov/clinvar) phenotype (supporting functional consequence of variants). Since *PTPN11* has no variants located in the three motifs, we did not verify the nature of the variants reported in *PTPN11*.

We found the respective amino acid index of the two SNV locations in gene *ACTB* using UCSC Genome Browser human GRCh37/hg19 assembly chromosome locations. We then examined these two coordinates (N12 and A22) and three potential SNVs using the Missense Variants to Protein Structure Analysis Web Suite (MISCAST) Variant Analysis Suite (miscast.broadinstitute.org).

Given that *GNAO1* was not found in MISCASST database, which could be due to the unavailability of the entire protein structure, we searched SWISS-MODEL (swissmodel.expasy.org) and used similar protein structure aligned with P09471 (UniProt ID) to deduce variant structural locations. To check for functional consequences of identified variants in *GNAO1*, we employed ClinVar to assess their potential clinical pathogenicity.

All steps mentioned above for identifying clinically significant variants of ID genes are shown in Fig. 1.

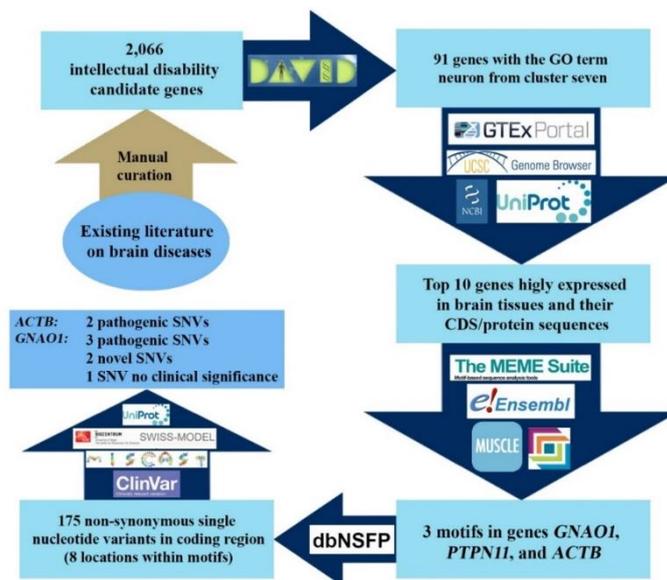


Fig. 1. Workflow of prioritizing coding region missense variants of genes associated with ID.

### 3. Results

#### 3.1. ID Related Gene Selection

To begin, our manual curation process for selecting ID candidate genes, which is discussed in more detail in our previous study on miRNA and mRNA pairs, resulted in a total of 2,066 genes from published literatures. The full list of genes is reported in [18] and also available upon request.

#### 3.2. GO Term Elucidation and Gene Expression Evaluation for ID Related Genes

All 2,066 genes were categorized into clusters based on their functionality and ordered by their enrichment score. Out of all clusters, cluster seven demonstrated the highest level of neuron engagement. After eliminating duplicates, we obtained a total of 91 genes.

For each selected gene, we assessed its expression levels in 54 different tissues as reported by GTEx, 13 of which are brain. Our results found ten genes, namely *PTPN11*, *GNAO1*, *SOX2*, *APBB1*, *STXBP1*, *TUBB2A*, *RAC1*, *KIF5C*, *GJA1* and *ACTB*, to be highly expressed in brain tissues when compared to the rest. We averaged the score of brain tissue expressions for all ten genes as shown in Fig. 2.

As seen, *ACTB* demonstrated extremely high brain tissue expressions. In fact, it exhibited the highest expression values in the cortex, frontal cortex, spinal cord (cervical c-1) and basal ganglia, which suggests a strong connection with brain and intellectual functions such as speech, thinking, and memory, and the possibility that dysfunctional *ACTB* proteins caused by mutations could lead to detrimental effects [19]. In addition, we inspected the expression graphs of each selected gene and found that *GNAO1* showed low expressions in other human tissues but abnormally high associations with cerebellum and cerebellar hemisphere, a part of the brain in charge of voluntary movements [20]. Analysis of past studies revealed

that cerebellar failure is evident in several developmental disorders, “including autism, ADHD, and developmental dyslexia,” and that cerebellar ataxia is often followed by ID later in childhood and adult life [21].

### 3.3. Protein Sequence Alignment and Conserved Domain Analysis

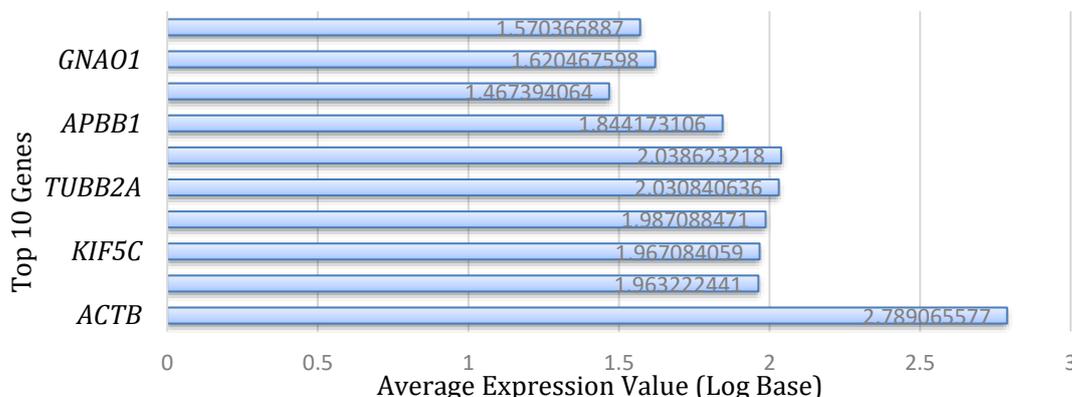


Fig. 2. Average expression of ten genes in 13 brain tissues (GTEx database).

A multiple sequence alignment on the protein sequences of the ten genes revealed many significantly conserved domains (distinct structural units responsible for protein interactions and functions). We investigated the functional abilities of these domains as well as PPI.

The results we obtained from NCBI Batch CD-Search tool of the nine genes revealed that *GNAO1* and *RAC1* belong to the same superfamily cl38936, the P-loop NTPase superfamily, characterized by the presence of the Walker A and Walker B motifs responsible for nucleotide phosphate binding [22]. More importantly, the domain search identified the proteins association with G-alpha proteins. Orphan G protein-coupled receptors (GPCR) proteins play important roles in causing neurological diseases and are potential targets for effective psychopharmacology [5].

The PPI data of the two genes *ACTB* and *GNAO1* obtained from STRING are shown in Fig. 3. The protein *ACTB* interacts with other proteins with similar molecular functions such as actin binding and actin cytoskeleton organization. When we changed the settings by increasing the number of interactors shown on the figure, we interestingly discovered *ACTB* protein to be interacting with *RAC1* protein with an interaction combined score of 0.950. The protein *GNAO1* interacts with other guanine nucleotide-binding proteins from the G-alpha group. Similarly, a change in settings to show more interactors revealed *GNAO1* to be interacting with *RAC1* with a combined score of 0.924 reported by STRING.

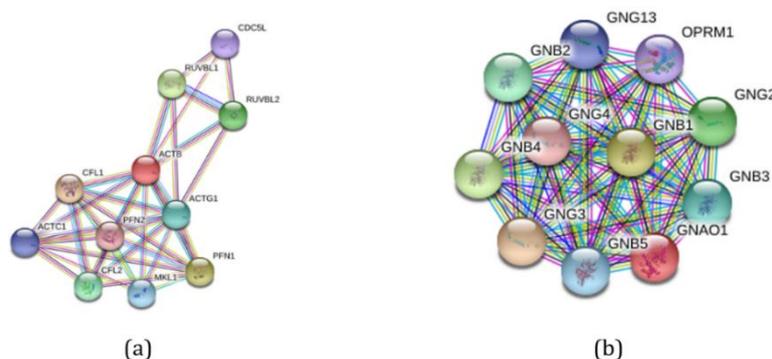


Fig. 3. Protein-protein interaction data of *ACTB* and *GNAO1* limited to top ten interactors. (a) PPI of *ACTB*. (b) PPI of *GNAO1*.

### 3.4. Motif Identification and Missense Variant Classification

Our results showed three significant motifs (Fig. 4) in the CDS sequences found in all nine genes (except *TUBB2A*), which highly suggests that these motifs are potentially associated with ID.

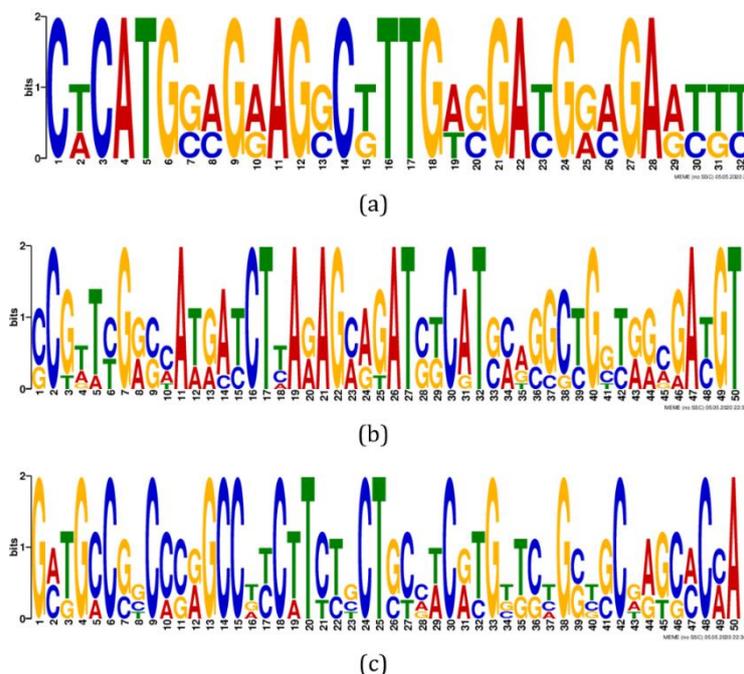


Fig. 4. Three most significant motifs in the CDS sequences of 9 genes. (a) Motif 1. (b) Motif 2. (c) Motif 3.

A total of 175 SNVs located in the three genes (*ACTB*, *GNAO1*, and *PTPN11*) were identified, among these eight possible variant coordinates are located within the three motifs. Though none of the SNVs reported by dbNSFP for gene *PTPN11* are located in the three most significant motifs, *PTPN11*, with other genes, does share the second-most significant motif with a p-value of 5.55e-18 reported by MEME.

On location chr7:5569255, which is in the *ACTB* gene, there reside two SNVs that are located in the most significant motif with a p-value of 2.48e-15. The two variants on index 12 are N>D and N>H, both of which were reported pathogenic and causative of Baraitser-Winter syndrome. Baraitser-Winter syndrome is a multiple congenital anomaly syndrome characterized by ID, thus suggesting the high possibility for this variant to also be associated with ID [23]. Similarly, the third SNV in *ACTB* located at chr7:5569225 with an amino acid index of 22 reported a single variant A>T and a connection with Baraitser-Winter syndrome. Therefore, all three possible variants in *ACTB* are pathogenic, and its association with the Baraitser-Winter syndrome suggests its role in causing ID.

The splice donor site variant (chr16:56226265) in *GNAO1* is identified as two potential SNVs (G>R and G>W) inside the first exon sequence (chr16:56226148-chr:56226266), in which the second motif is located. Clinical significance on ClinVar reported both SNVs as pathogenic and associated with early infantile epileptic encephalopathy (EIEE1), a type of neurological disorder caused by brain malformation or genetic mutations due to the lack of *ARX* proteins which disrupts normal brain development and leads to seizures and ID [24]. Our results verified the strong correlation between *GNAO1* variant mutations and EIEE1 that leads to ID later in life.

Furthermore, a group of five locations in *GNAO1*, specifically chr16:56370698, chr16:56370711, chr16:56370729, chr16:56370732, and chr16:56370741, are all located in the third most significant motif. The third most significant motif, also shared by gene *ACTB*, has a p-value of 1.78e-18 in gene *GNAO1*. The

fact that the five SNVs are located close to each other but not consecutive further suggests the potential influence through miRNA binding. Clinical reports on ClinVar also proved two of the five SNVs-chr16:56370732 and chr16:56370741-to be probably pathogenic and pathogenic respectively. Out of the remaining three variants, two were novel variants we discovered at chr16:56370698 and chr16:56370711, and although the last SNV at chr16:56370729 did not report any clinical significance, we believe all three SNVs are highly likely to be pathogenic and strongly associated with ID and other related syndromes as were the other SNVs we identified. In the future, clinical experiments could seek to verify our results. The Table 1 below summarizes our discovery and the ten SNVs we analyzed in our study.

Table 1. Non-Synonymous Variants in *ACTB* and *GNAO1* Obtained from dbNSFP

| chr | pos (hg19) | ref | alt | aaref | aaalt | Gene         | Location             | Motif | Clinical Significance              |
|-----|------------|-----|-----|-------|-------|--------------|----------------------|-------|------------------------------------|
| 7   | 5569225    | C   | T   | A     | T     | <i>ACTB</i>  | CDS                  | 1     | Pathogenic                         |
| 7   | 5569255    | T   | C   | N     | D     | <i>ACTB</i>  | CDS                  | 1,3   | Pathogenic                         |
| 7   | 5569255    | T   | G   | N     | H     | <i>ACTB</i>  | CDS                  | 1,3   | Pathogenic                         |
| 16  | 56226265   | G   | C   | G     | R     | <i>GNAO1</i> | Splice donor site    | 2     | Likely pathogenic                  |
| 16  | 56226265   | G   | T   | G     | W     | <i>GNAO1</i> | Splice donor site    | 2     | Likely pathogenic                  |
| 16  | 56370698   | G   | A   | E     | K     | <i>GNAO1</i> | Splice acceptor site | 3     | Novel variant                      |
| 16  | 56370711   | C   | A   | A     | D     | <i>GNAO1</i> | CDS                  | 3     | Novel variant                      |
| 16  | 56370729   | C   | T   | A     | V     | <i>GNAO1</i> | CDS                  | 3     | Clinical significance not reported |
| 16  | 56370732   | T   | C   | L     | P     | <i>GNAO1</i> | CDS                  | 3     | Likely pathogenic                  |
| 16  | 56370741   | A   | G   | Y     | C     | <i>GNAO1</i> | CDS                  | 3     | Likely pathogenic                  |

### 3.5. Variant 3D Structural Location

Fig. 5 showcases variants as described in Table 1 for *ACTB*. Although the entire protein structure of *GNAO1* is still unavailable for use, we obtained 3D figures from SWISS-MODEL that aligned it with similar protein structures. Fig. 6 showcases closed-up images of variants from *GNAO1*.

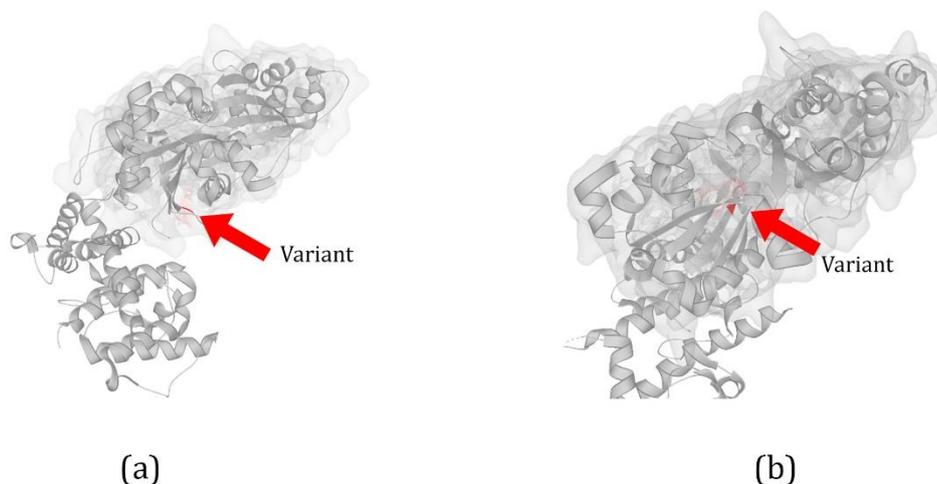


Fig. 5. SNVs Structural location for variants in *ACTB* (red highlighted). (a)chr7:5569225 N12 (b)chr7:5569255 A22.

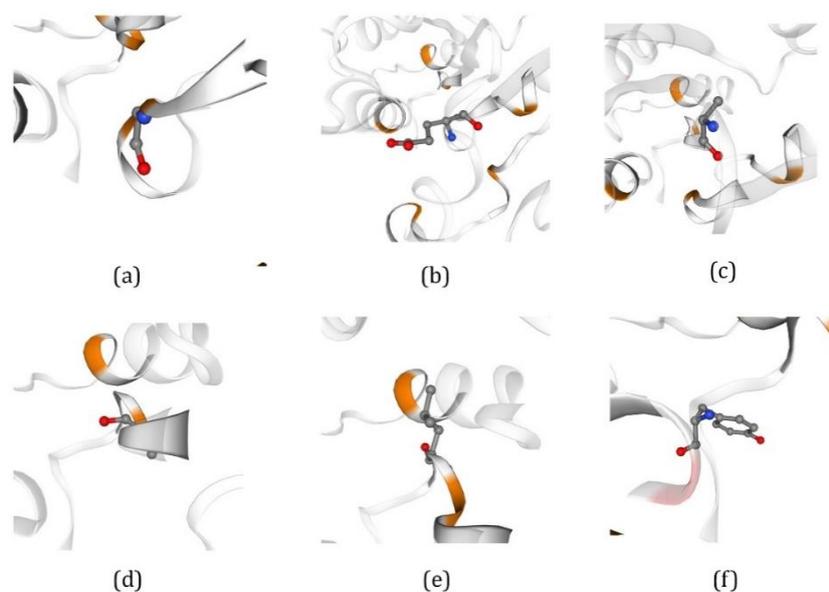


Fig. 6. SNVs structural location in *GNAO1* (blue and red nodes). (a)chr16:5622625 G40 (b)chr16:56370698 E217 (c)chr16:56370711 A221 (d)chr16:56370729 A227 (e)chr16:56370732 L228 (f)chr16:56370741 Y231.

#### 4. Discussion

The gene *ACTB* encodes a protein called beta ( $\beta$ )-actin, which is part of the actin protein family responsible for making up the structural framework inside cells. Mutations inside of *ACTB* have been found to cause Baraitser-Winter syndrome, a condition that affects brain and facial feature development [25]. These mutations change the amino acid sequence in  $\beta$ -actin leading to malformed cytoskeleton and cell structure, which ultimately affects the cell movement. Given that  $\beta$ -actin is a common cell present throughout one's body, dysfunctional  $\beta$ -actin proteins are major contributors to developmental issues, including brain growth [26].

*GNAO1* encodes proteins that represent the alpha subunit of the G-protein signal-transducing complex. Mutations within *GNAO1* have been found to cause neurodevelopmental disorder with involuntary movements (NEDIM) and EIEE1, both of which are severe neurologic disorders within the brain [27]. In fact, the two novel SNVs we discovered at chr16:56370698 and chr16:56370711 are highly likely to cause similar brain malfunction given the high expression levels of *GNAO1* in 13 brain tissues (Fig. 7). To validate our results through whole-genome sequencing (mainly wet-lab experiments) would require lots of additional resources and significant amount of time, but we definitely look to verify our discoveries through clinical trials in future research project. Nevertheless, through an extensive literature review on papers found in reputable academic databases such as PubMed, Google Scholar, and ScienceDirect, numerous existing studies that employed clinical trials with ID patients have verified the pathogenic role of *GNAO1* mutations in neurodevelopmental disorders [27]-[29]. In addition, we checked our identified variants against 1000 Genome data ([www.internationalgenome.org/about#1000G\\_PROJECT](http://www.internationalgenome.org/about#1000G_PROJECT)), a sequencing-based variant calling database with the largest public catalogue of human variation and genotype data, and found that the dbNSFP database we used in the study correctly reported *all* potential SNVs in the three genes.

The three genes (*ACTB*, *GNAO1*, and *PTPN11*) focused on in our study demonstrated a close relationship on proteomic levels as well, implying their similar biological functions. In the phylogenetic tree (available upon request), *PTPN11*, *ACTB*, and *GNAO1* were clustered closely together, which suggest that they share a closer relationship with each other than the remaining seven genes. Moreover, in another separate study by



## Author Contributions

Anna Chang Liu and Yongsheng Bai conducted the research; Anna Chang Liu, Junmeng Yang, Tina Yuan and Yongsheng Bai analyzed the data; Anna Chang Liu and Yongsheng Bai wrote the paper; all authors have approved the final version.

## Acknowledgment

The authors would like to thank Xiaoming Liu, Chunlei Wu, Chang Li, and Eric Boerwinkle for the dbNSFP database.

## References

- [1] Salvador, C., L., Reed, G. M., Vaez, A. L. M., Cooper, S. A., Martinez, L. R., Bertelli, M., *et al.* (2011). Intellectual developmental disorders: Towards a new name, definition and framework for "mental retardation/intellectual disability" in ICD-11. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 10(3), 175-180.
- [2] American Psychological Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*. Washington: American Psychiatric Publishing.
- [3] Harbin Clinic. Genetic brain disorder. Retrieved from <https://harbinclinic.com/term/genetic-brain-disorder>
- [4] Alavi, M. S., Shamsizadeh, A., Azhdari, Z. H., & Roohbakhsh, A. (2018). Orphan G protein-coupled receptors: The role in CNS disorders. *Biomed. Pharmacother*, 98, 222-232.
- [5] Khan, M. Z., & He, L. (2017). Neuro-psychopharmacological perspective of Orphan receptors of rhodopsin (class A) family of G protein-coupled receptors. *Psychopharmacology (Berl)*, 234(8), 1181-1207.
- [6] Gilissen, C., Hehir, K. J. Y., Thung, D. T., Vorst, V. D. M., Bon, V. B. W., Willemsen, M. H., *et al.* (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature*, 511(7509), 344-347.
- [7] Ligt, D. J., Willemsen, M. H., Bon, V. B. W., Kleefstra, T., Yntema, H. G., Kroes, T., *et al.* (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.*, 367(20), 1921-1929.
- [8] Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Ende, S., Schwarzmayr, T., *et al.* (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: An exome sequencing study. *Lancet*, 380(9854), 1674-1682.
- [9] Large-scale discovery of novel genetic causes of developmental disorders. *Nature*, 519(7542), 223-228. Retrieved December 24, 2014, from <https://www.nature.com/articles/nature14135>
- [10] Mahmoudi, E., & Cairns, M. J. (2017). MiR-137: An important player in neural development and neoplastic transformation. *Mol. Psychiatry*, 22(1), 44-55.
- [11] Asim, A., Kumar, A., Muthuswamy, S., Jain, S., & Agarwal, S. (2015). Down syndrome: An insight of the disease. *Journal of Biomedical Science*, 22(1), 41-41.
- [12] Chai, M., Su, L., Hao, X., Zhang, M., Zheng, L., Bi, J., *et al.* (2019). Identification of a thymus microRNA-mRNA regulatory network in Down syndrome. *Molecular Medicine Reports*, 20(3), 2063-2072.
- [13] Franzoni, E., Booker, S. A., Parthasarathy, S., Rehfeld, F., Grosser, S., Srivatsa, S., *et al.* (2015). miR-128 regulates neuronal migration, outgrowth and intrinsic excitability via the intellectual disability gene Phf6. *eLife*, 4, e04263.
- [14] Lim, J. H., Han, Y. J., Kim, H. J., Kim, M. Y., Park, S. Y., Cho, Y.-H., *et al.* (2018). Integrative analyses of genes and microRNA expressions in human trisomy 21 placentas. *BMC Medical Genomics*, 11(1), 46.

- [15] Siew, W. H., Tan, K. L., Abbaspour B. M., Cheah, P. S., & Ling, K. H. (2013). MicroRNAs and intellectual disability (ID) in down syndrome, X-linked ID, and fragile X syndrome. *Frontiers in Cellular Neuroscience*, 7(41), Article 4.
- [16] Zablotskaya, A., Van, E. H., Verstrepen, K. J., Froyen, G., & Vermeesch, J. R. (2018). Mapping the landscape of tandem repeat variability by targeted long read single molecule sequencing in familial X-linked intellectual disability. *BMC Medical Genomics*, 11(1), 123.
- [17] Liu, X., Wu, C., Li, C., & Boerwinkle, E. (2016). dbNSFP v3.0: A one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum. Mutat.*, 37(3), 235-241.
- [18] Junmeng, Y., Anna, L., Isabella, H., & Yongsheng, B. (2020). Bioinformatics analysis revealed novel 3'UTR variants associated WITH intellectual disability. *Genes*, 11(9), 998.
- [19] Riva, D., Taddei, M., & Bulgheroni, S. (2018). The neuropsychology of basal ganglia. *Eur. J. Paediatr. Neurol.*, 22(2), 321-326.
- [20] Stoodley, C. J., & Schmahmann, J. D. (2018). Functional topography of the human cerebellum. *Handb. Clin. Neurol.*, 154, 59-70.
- [21] Stoodley, C. J. (2016). The cerebellum and neurodevelopmental disorders. *Cerebellum*, 15(1), 34-37.
- [22] Romero, M. L., Yang, F., Lin, Y. R., Toth, P. A., Berezovsky, I. N., Goncarenco, A., *et al.* (2018). Simple yet functional phosphate-loop proteins. *Proc. Natl. Acad. Sci. USA*, 115(51), e11943-e11950.
- [23] Yates, T. M., Turner, C. L., Firth, H. V., Berg, J., & Pilz, D. T. (2017). Baraitser-winter cerebrofrontofacial syndrome. *Clin. Genet.*, 92(1), 3-9.
- [24] Medline Plus, U.S. National Library of Medicine (2020). Early infantile epileptic encephalopathy 1. Retrieved from <https://ghr.nlm.nih.gov/condition/early-infantile-epileptic-encephalopathy-1#genes>
- [25] Cuvertino, S., Stuart, H. M., Chandler, K. E., Roberts, N. A., Armstrong, R., Bernardini, L., *et al.* (2017). ACTB loss-of-function mutations result in a pleiotropic developmental disorder. *Am. J. Hum. Genet.*, 101(6), 1021-1033.
- [26] Dominguez, R., & Holmes, K. C. (2011). Actin structure and function. *Annu. Rev. Biophys.*, 40, 169-186.
- [27] Marcé, G. A., Dalton, J., López, P. J., García, J. M. C., Monge, G. L., Cuenca, L. E., *et al.* (2016). GNAO1 encephalopathy: Further delineation of a severe neurodevelopmental syndrome affecting females. *Orphanet Journal of Rare Diseases*, 11(1), 38.
- [28] Danti, F. R., Galosi, S., Romani, M., Montomoli, M., Carss, K. J., Raymond, F. L., *et al.* (2017). GNAO1 encephalopathy: Broadening the phenotype and evaluating treatment and outcome. *Neurology Genetics*, 3(2), e143.
- [29] Nakamura, K., Hirofumi, K., Tenpei, A., Shiina, M., Kato, M., Hoshino, H., *et al.* (2013). De Novo mutations in GNAO1, encoding a Gα subunit of heterotrimeric G proteins, cause epileptic encephalopathy. *The American Journal of Human Genetics*, 93(3), 496-505.
- [30] Hingorani, A. D., Kuan, V., Finan, C., Kruger, F. A., Gaulton, A., Chopade, S., *et al.* (2019). Improving the odds of drug development success through human genomics: modelling study. *Scientific Reports*, 9(1), 18911.
- [31] Nelson, M. R., Tipney, H., Painter, J. L., Shen, J., Nicoletti, P., Shen, Y., *et al.* (2015). The support of human genetic evidence for approved drug indications. *Nat. Genet.*, 47(8), 856-860.
- [32] Finan, C., Gaulton, A., Kruger, F. A., Lumbers, R. T., Shah, T., Engmann, J., *et al.* (2017). The druggable genome and support for target identification and validation in drug development. *Science Translational Medicine*, 9(383), eaag1166.



**Anna Chang Liu** was born in Zhengzhou, China. She studies business administration and computer science at Appleby College in Ontario, Canada. Her previous works have been published and presented at the International Conference on Intelligent Biology and Medicine 2020 and Conference on Intelligent Systems of Molecular Biology 2020. Liu's main research subjects include bioinformatics, computer science, and environmental studies.